

**SUBMISSION AGREEMENT
BETWEEN
THE CLIMATE DATABASE MODERNIZATION PROGRAM CONTRACTOR (IBM)
AND
THE NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION
FOR IMAGED WEATHER AND CLIMATE DOCUMENTS FROM EDADS**

2011-08-09

Introduction

This document represents the agreement that the Climate Database Modernization Program Contractor (IBM) (CDMP) (the "Provider") and the National Centers for Environmental Information (NCEI) (the "Archive") have reached for submitting the Provider's data, Imaged Weather and Climate Documents from EDADS, to the Archive for long-term preservation. It represents a joint effort between the Provider and the Archive to accurately document the agreement and the expectations between the two groups.

In order to ensure that the quality and integrity of the archived data is not compromised, the Provider and the Archive agree to maintain this agreement with accurate and up-to-date information through the life of the data submission.

Add comments as needed

Contacts

Persons included in all communications regarding the data submission.

Provider Contacts

Point of Contact	Main Contact
William Angel	Cynthia Karl
NCDC	NCDC
Program Manager	+1 828-271-4370
+1 828-271-4459	cynthia.b.karl@noaa.gov
william.e.angel@noaa.gov	

Main Contact for descriptive information

Jason Cooper
NCDC
+1 828-271-4382
jason.cooper@noaa.gov

Archive Contacts

Data Acquisition

Heather Anderson

NCDC

Archive Branch

828-271-4863

heather.anderson@noaa.gov

David Bowman

NCDC/RSAD/AB

IT Specialist

+1 828-271-4368

david.p.bowman@noaa.gov

Tammy Scott

NCDC/RSAD/AB

IT Specialist

+1 828-271-4175

tammy.scott@noaa.gov

Data Acquisition

Philip Jones

NCDC/RSAD/AB

Archive Branch

+1 828-271-4472

philip.jones@noaa.gov

Data Overview

This is a collection of libraries containing scientific observation documents (in various formats) and their associated metadata/indexes. This is the repository for the documents processed by the Climate Database Modernization Project. These data are in situ data from the military, NOAA, states, DOA, and other related climate related data.

The data was distributed through CDMP EDADS (Environmental Document Access and Display System) software. A Sharepoint database which was located off-site at the providers facility.

These data are being transferred back to NCDC by staggering multiple drives of 3-2 TB drives each, totaling 15 TB. Once each drive is delivered to NCDC these data will be run through validation and security checked, then copied on to the Archive server.

Once archived depending on the dataset, the collections will be copied onto a local server to be distributed through a new controlled access and display system, EV2, which will interact with IPS giving limited public access, or transferred by to the data owner.

Applicable and Reference Documents

Documents applicable to or referenced from this agreement.

1. Citation for applicable/reference document, e.g., an archive appraisal or project requirements document

Submission Scope

Active Submission Period

- 2011-09-30

Data Types

Below is a summary of the data sizing and submission schedule by data type group. Enter information on at least one data type.

Data Type Name	Data Sizing	Submission Schedule
----------------	-------------	---------------------

Initial Delivery	3TB	High priority data collections to be sent by IBM to NCDC by 08/31/2011; remaining by 10/31/2011
------------------	-----	---

Reviews and Testing

These data collections were originally quality controlled at the contractor's facility

These data are being transferred to NCDC by staggering multiple drives of 3-2 TB drives each, totaling 15 TB. Once each drive is delivered to NCDC they will be run through validation and security checks, then copied on to the Archive server.

Once archived depending on the data set, the collections will be copied onto a local server to be distributed through a new controlled access and display system, EV2, which will interact with IPS giving limited public access, or transferred by to the data owner.

Providing System

Identification of the system providing the data to NCEI.

System Name: gulp2
System Owner: NCDC
Physical Location: NCDC, Asheville, NC
Additional Information: Add comments as needed on applicable data types, etc.

Transfer Interface

The provider will FTP push all files to a directory on an Archive server as specified by the Archive. The Provider will be responsible for verifying the success of the FTP session.

Submission File Inventory

Information on each submitted file type from the Provider. Information on multiple file types can be added below.

File Type Name: See Appendix A for all 130 files

File Name Pattern:

See Appendix A

File Name Field Definitions:

See Appendix A

Example File Name:

See Appendix A

File Format: TIFF

File Compression: None

File Size Average: 15GB

File Count (Rate): minimum of 58,445,454 files

Data Volume (Rate): minimum of 15 TB total, delivered in multiple ~2-3TB disks

Submission Schedule: High priority data to be sent by IBM to NCDC by 08/31/2011; remaining by 10/31/2011

Additional Information: Multiple data formats, see Appendix A.

Descriptive Information Attributes:

None: date/time fields are usually used as descriptive attributes for this file.

Submission Manifest

A submission manifest file with a 32-character MD5 checksum value is required for each submitted file in order to ensure the integrity of the submitted data.

File Content Specification:

Define the contents of the submission manifest, e.g., "A submission manifest file contains a comma delimited list of submitted file names and associated checksums for one or more submitted files."

File Transmission:

The frequency and mechanism of a submission manifest transmission.

File Name Pattern:

Naming pattern used to identify the submission manifest file

File Name Definitions:

Definitions of the fields in the submission manifest file name pattern

Example File Name:

An actual submission manifest file name

Archive Ingest

Ingest processing steps at the Archive and communication with the Provider.

Receipt Verification:

The Archive will use the provided file name and 32-character MD5 checksum value to verify the integrity of a delivered file. The Archive will also manually check the received files for any file packaging errors.

Error Reconciliation:

The Archive will report any problems or errors with file integrity, file name, checksum validation, or other errors that inhibit the data ingest and archive to the Provider. A new corresponding submission manifest will be required for files re-submitted by the Provider.

Receipt Confirmation:

The Archive will report any problems or errors with file integrity, file name, checksum validation, or other errors that inhibit the data ingest and archive to the provider via email. A new corresponding submission manifest will be required for files re-submitted by the provider.

Quality Assurance:

The Archive will review sample data and documentation as provided before data submission. No reviews of the data by the Archive are required once submitted.

Archive File Packaging:

See Appendix A for a description of file packaging or re-naming by the Archive upon ingest.

Archive Storage

Archive attributes of each archived file type.

Archive File Type Name: Descriptive name for this archive file type	
Archive File Attributes/IDs:	
Attribute/ID Type	Value
Storage System	NCDC HDSS
Archive File ID	/aab/xxxx/39<xx_xx>/<directory name>_img_<startofimagerange>_<endofimagerange>_r<dateoflastrevisionor delivery>.<format>
Data Family ID	TBD
Data Type ID	Attribute/ID value
ACMI CLASS ID	Attribute/ID value
ACMI File	/ INSITU /
Metadata ID	/ INSITU /
Additional Information:	Attribute/ID value

Archive Updates

Data submissions intended to update an existing archive record require adequate notification and justification. Updates can supersede previous data submissions as a newer or improved version, however any previously submitted data will not be removed from the archive for the purpose maintaining version control and traceability in the archive.

Retention Schedule

The data will be retained in the Archive for long-term preservation in accordance with NOAA data management standards. Information on data usage and archive value may be used for making decisions on continuing the duration of the archive.

(Notional) Disposition: Unknown/TBD

Constraints

Constraint Type	Description
Access	1. Not all the libraries belong to NCDC or even NOAA. 2. The Station History forms many contain sensitive information.

User Community

Description of the user groups and how the archived data will likely be used.

User Documentation and Metadata

The Provider will supply information to the Archive for writing and maintaining standard archive metadata, which includes data discovery information, references and data archive access links for users. The following published documents and archived items will be referenced from the metadata and made available to users.

Representation Information Items

For data to be useful to users, present and future, its format specification and characteristics must be documented and preserved with the data. Representation Information provides users with syntax (structure) and/or semantics (meaning) to decode the encoded data.

Item	Description
Item name or citation	Item description or intended use

Preservation Descriptive Information Items

Preservation Descriptive Information items contain context, provenance, and/or quality information for the data.

Item	Description
Item name or citation	Item description or intended use

Access and Dissemination

The Archive will provide access services for the data and supporting information to the designated user community.

Additional Terms

None.